www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Better observations, better forecasts?

F. Massonnet

O. Bellprat, F. J. Doblas-Reyes

ECMWF
29th February 2016

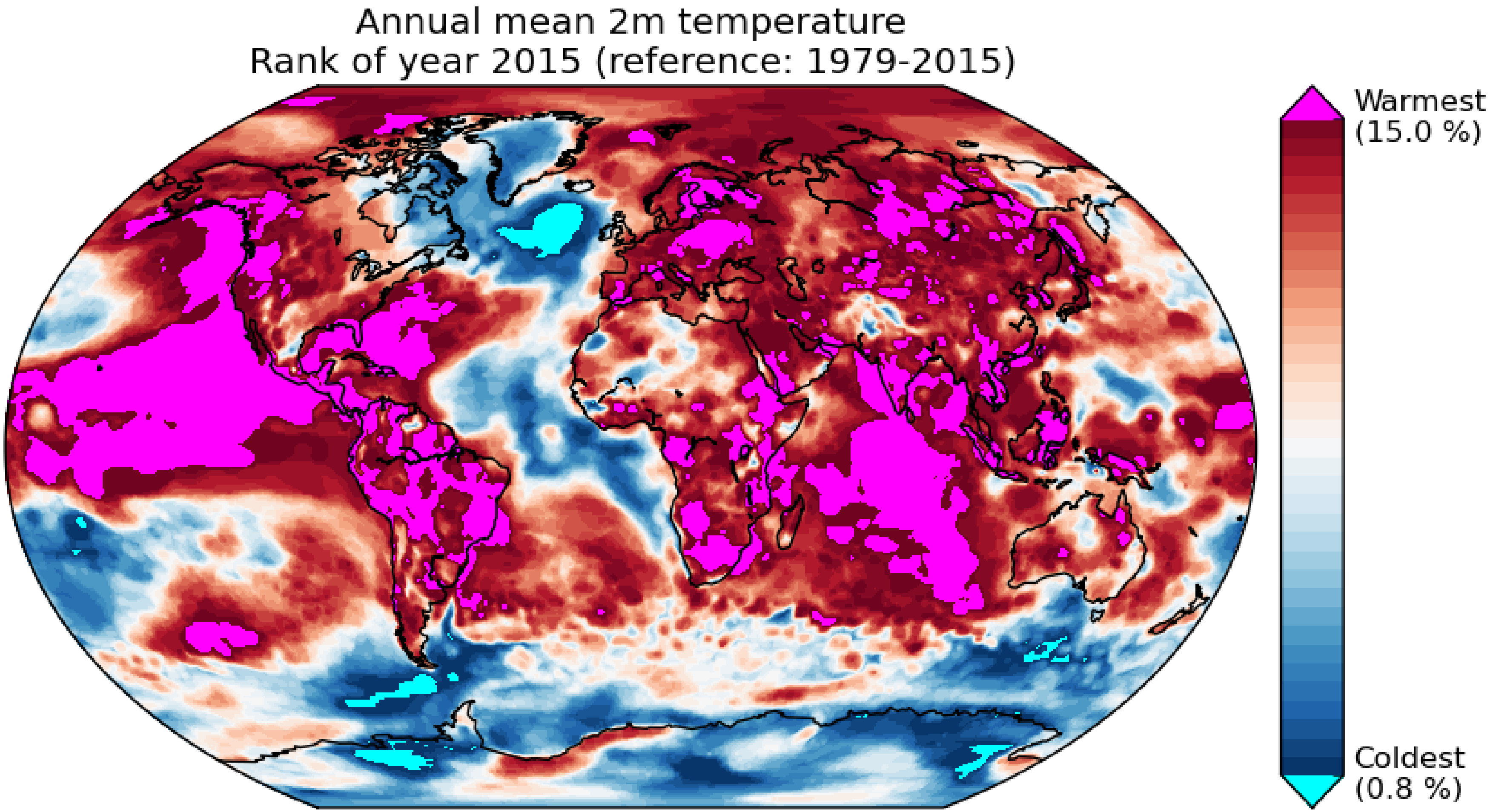
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

SPECS
Seasonal-to-decadal climate Prediction for the improvement of European Climate Services

GOBIERNO DE ESPAÑA

MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Our climate is changing rapidly



Data: ERA-Interim

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

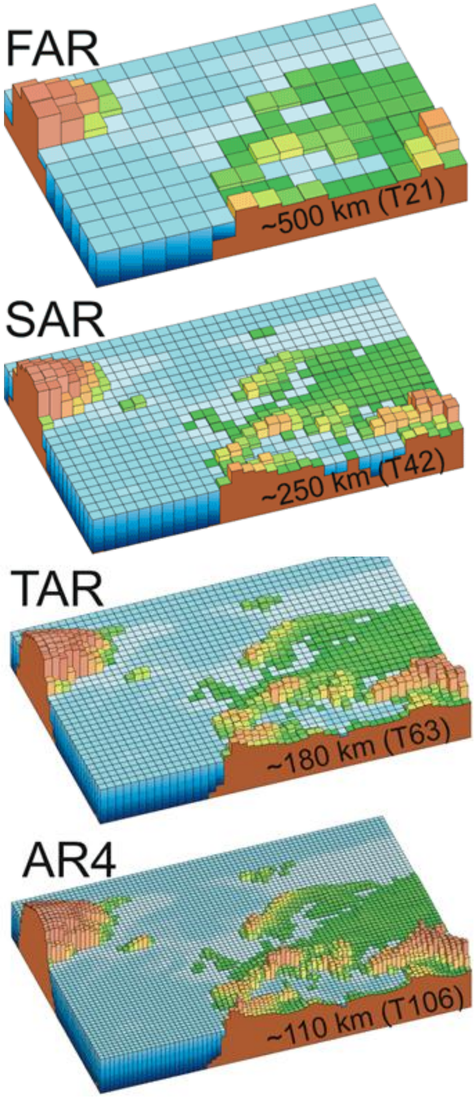
The need for high-quality observations has never been as pressing as it is today

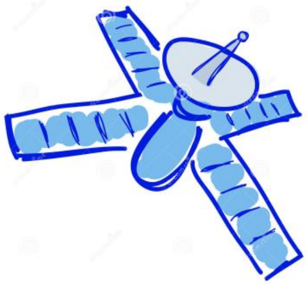
Climate is changing rapidly
but parts of our planet remain largely unexplored

Internal variability is large and underlying mechanisms are not completely understood
e.g. Deser et al., *J. Clim*, 2015

Model predictions/projections remain uncertain
Reliable observations are crucial for model evaluation

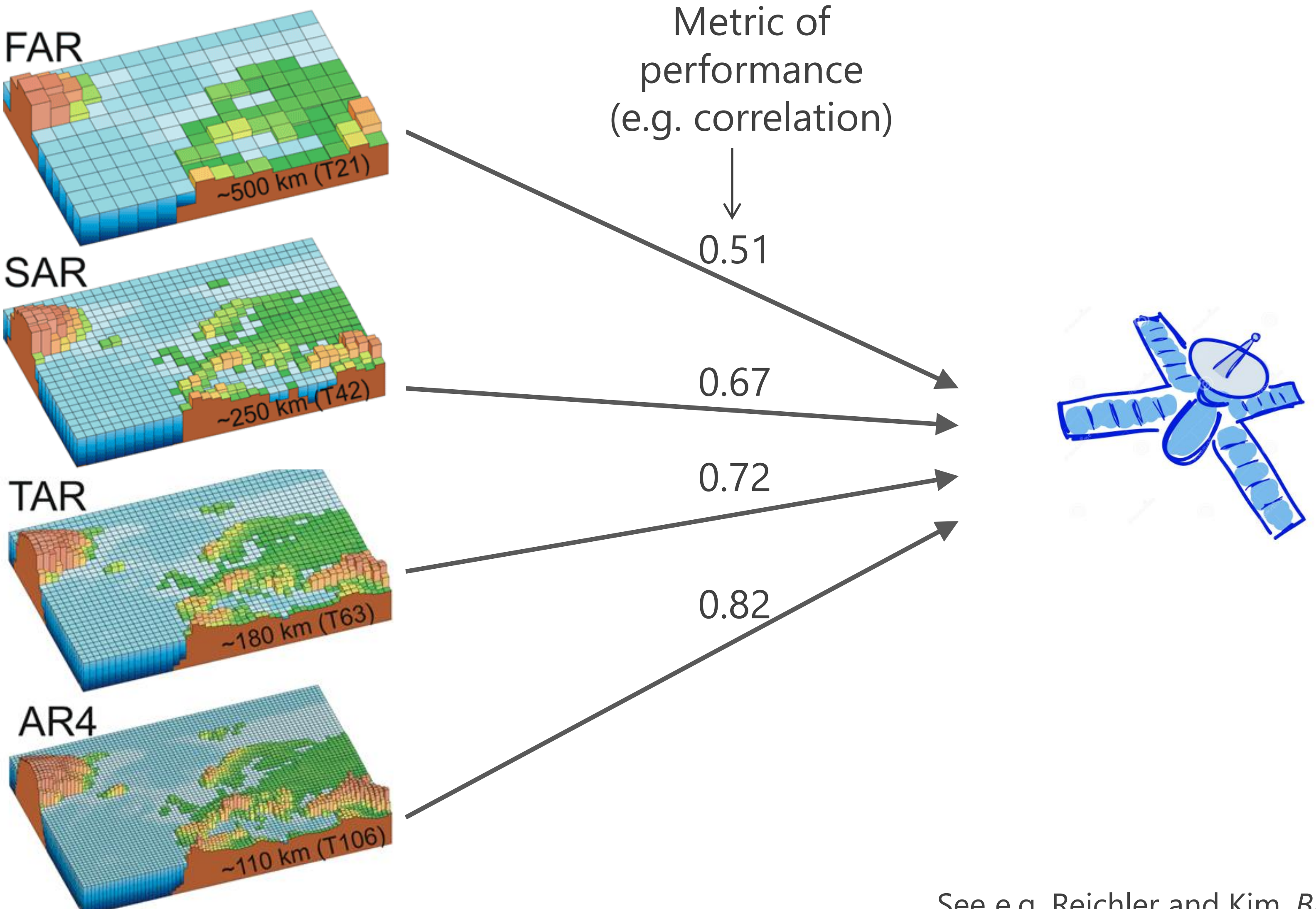
The classical approach of evaluation: several models for one observation





http://eo.ucar.edu/staff/rrussell/climate/modeling/climate_model_resolution.html

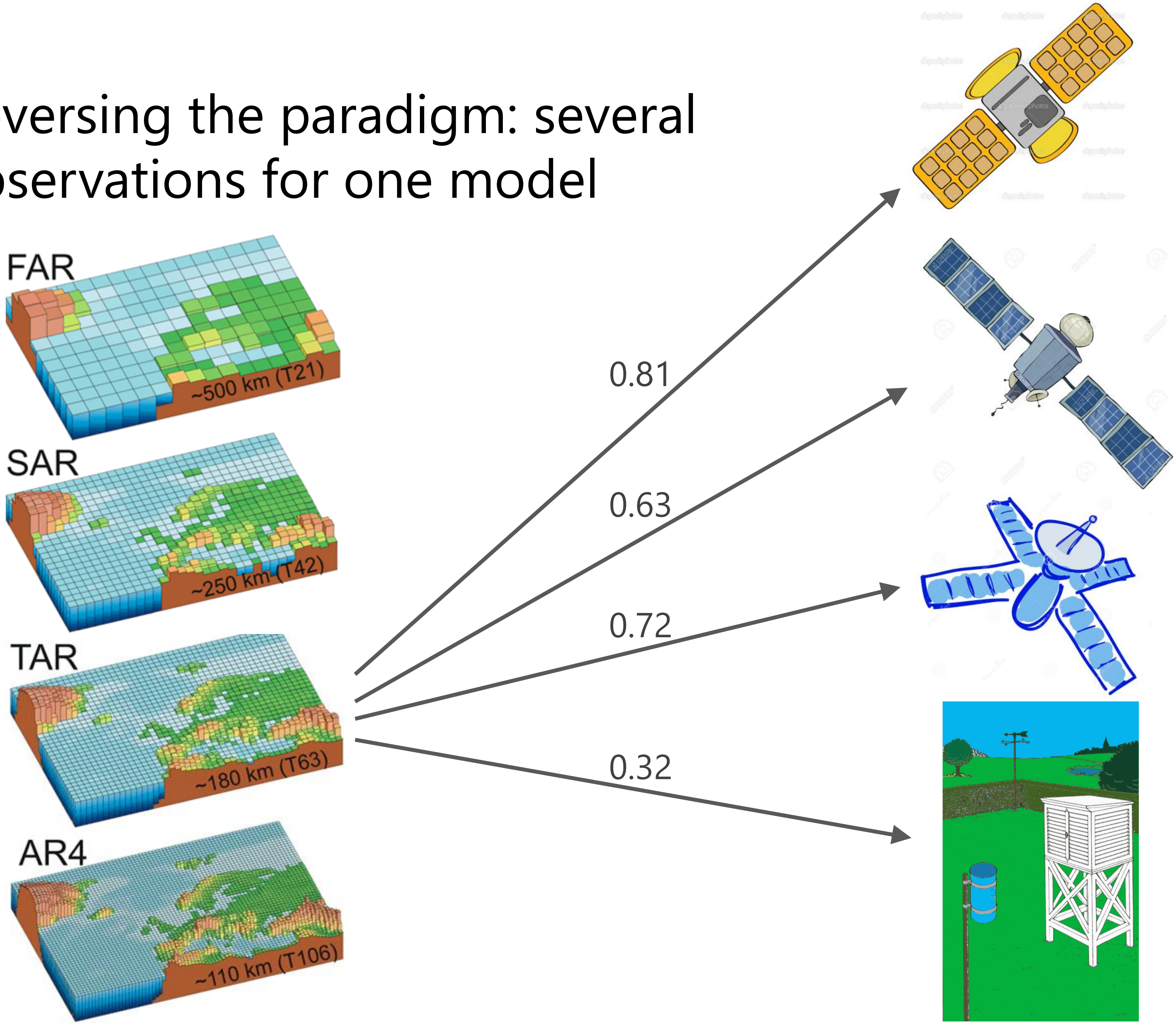
The classical approach of evaluation: several models for one observation



http://eo.ucar.edu/staff/rrussell/climate/modeling/climate_model_resolution.html

See e.g. Reichler and Kim, *BAMS*, 2008

Reversing the paradigm: several observations for one model



http://eo.ucar.edu/staff/rrussell/climate/modeling/climate_model_resolution.html

« Better observations yield better forecast verification scores »

1. Intuition ▸ 2. Formalization ▸ 3. Confirmation

« Better observations yield better forecast verification scores »

1. Intuition ▸ 2. Formalization ▸ 3. Confirmation

If metrics of performance (e.g., correlation, RMSE) are appropriate tools to reflect the quality of a modelling/forecast system,

Reichler and Kim, *BAMS*, 2008 (CMIP3→CMIP5)
Scaife et al., *GRL*, 2014 (NAO; sampling)
Massonnet et al., *The Cryosph*., 2012 (sea ice)
Msadek et al., *GRL*, 2014 (sea ice)

then, they can also reveal the underlying quality of an observational dataset.

This is because metrics of performance are symmetric from a mathematical point of view.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

« Better observations yield better forecast verification scores »

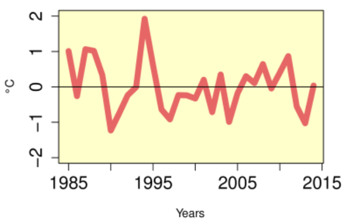
1. Intuition ▸ **2. Formalization** ▸ 3. Confirmation

A signal-plus-noise toy model

Interannual variability

TRUTH $X_t = \boxed{\epsilon}$

$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon)$$

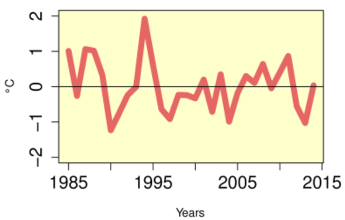


[Model from Weigel et al., *QJRMS*, 2008. See also Siegert et al., 2015]

A signal-plus-noise toy model

Interannual variability

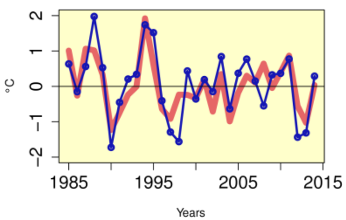
TRUTH $X_t = \boxed{\epsilon}$ $\qquad \epsilon \sim \mathcal{N}(0, \sigma_\epsilon)$



Interannual variability

Measurement and representativity error

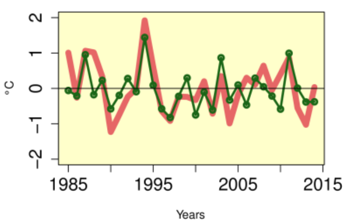
OBS $X_o = \boxed{\epsilon} + \boxed{\eta_o}$ $\qquad \eta_o \sim \mathcal{N}(0, \sigma_o)$



Interannual variability

Model forecast error (physics, initial conditions, resolution) + irreducible error (atmosphere)

MODEL $X_f = \alpha\boxed{\epsilon} + \boxed{\eta_f + \eta_m}$ $\qquad \eta_f \sim \mathcal{N}(0, \sigma_f), \qquad \eta_m \sim \mathcal{N}(0, \sigma_m)$



All error terms are assumed to be uncorrelated

[Model from Weigel et al., *QJRMS*, 2008. See also Siegert et al., 2015]

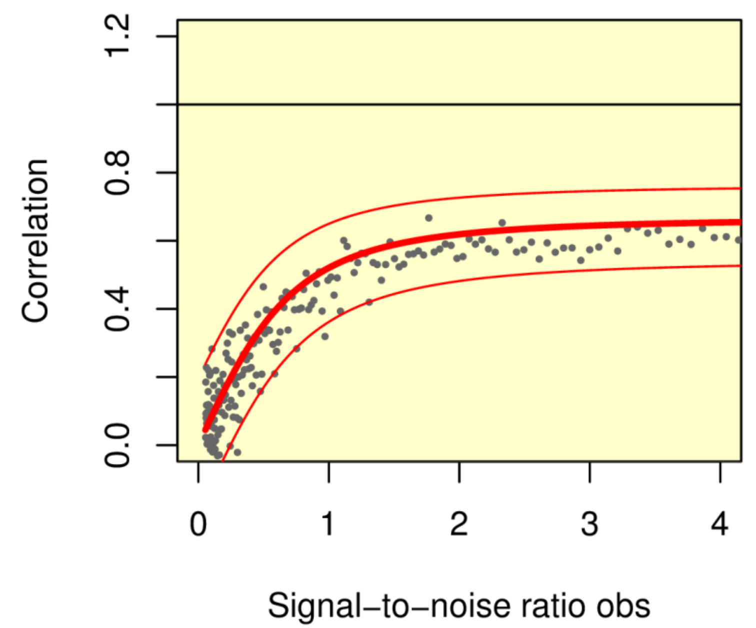
In this very simple paradigm, model and observational errors play interchangeable roles

$$\rho(X_o, X_f) = \frac{1}{\sqrt{\left(1 + \frac{\sigma_o^2}{\sigma_\epsilon^2}\right) \cdot \left(1 + \frac{(\sigma_f^2 + \sigma_m^2)/\alpha^2}{\sigma_\epsilon^2}\right)}}$$

Correlation **increases** when

- Model explains more variability,
- Model error decreases,
- Climate signal is stronger,
- Observational error decreases.

If error statistics are known, the dependence can be predicted



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

« Better observations yield better forecast verification scores »

1. Intuition ▸ 2. Formalization ▸ **3. Confirmation**

Seasonal forecasts of summer sea surface temperature

CanCM3/4, MPI-ESM-L/M/HR, CNRM, EC-EARTH (3 versions)

11 models

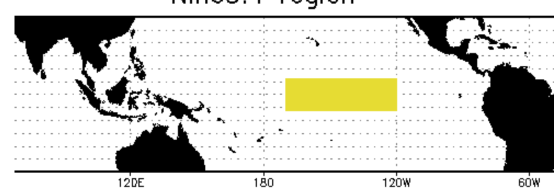
10 members

4-month forecasts initialized in May

440 forecasts (<u>not</u> independent from each other)

Period for evaluation: 1993-2009

Nino3.4 region



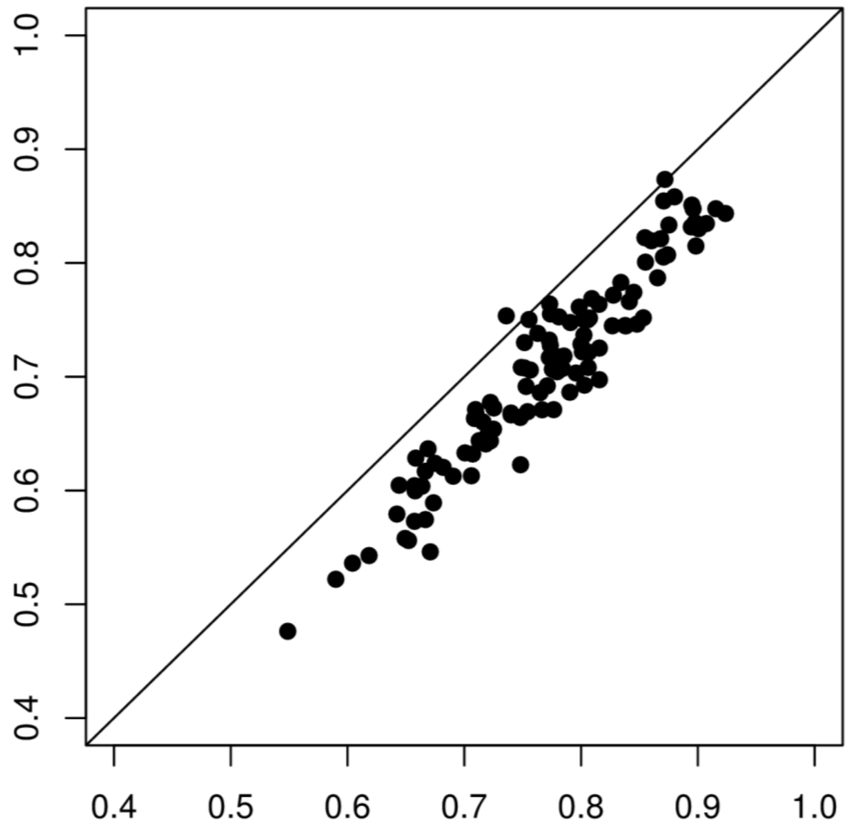
ESA-CCI ERA-Int ERSST HadISST ERSST4

5 observational datasets for verification

(<u>not</u> independent from each other either)

The choice of the validation product has a systematic impact on correlations

Skill of August forecast of SST (1993-2009)

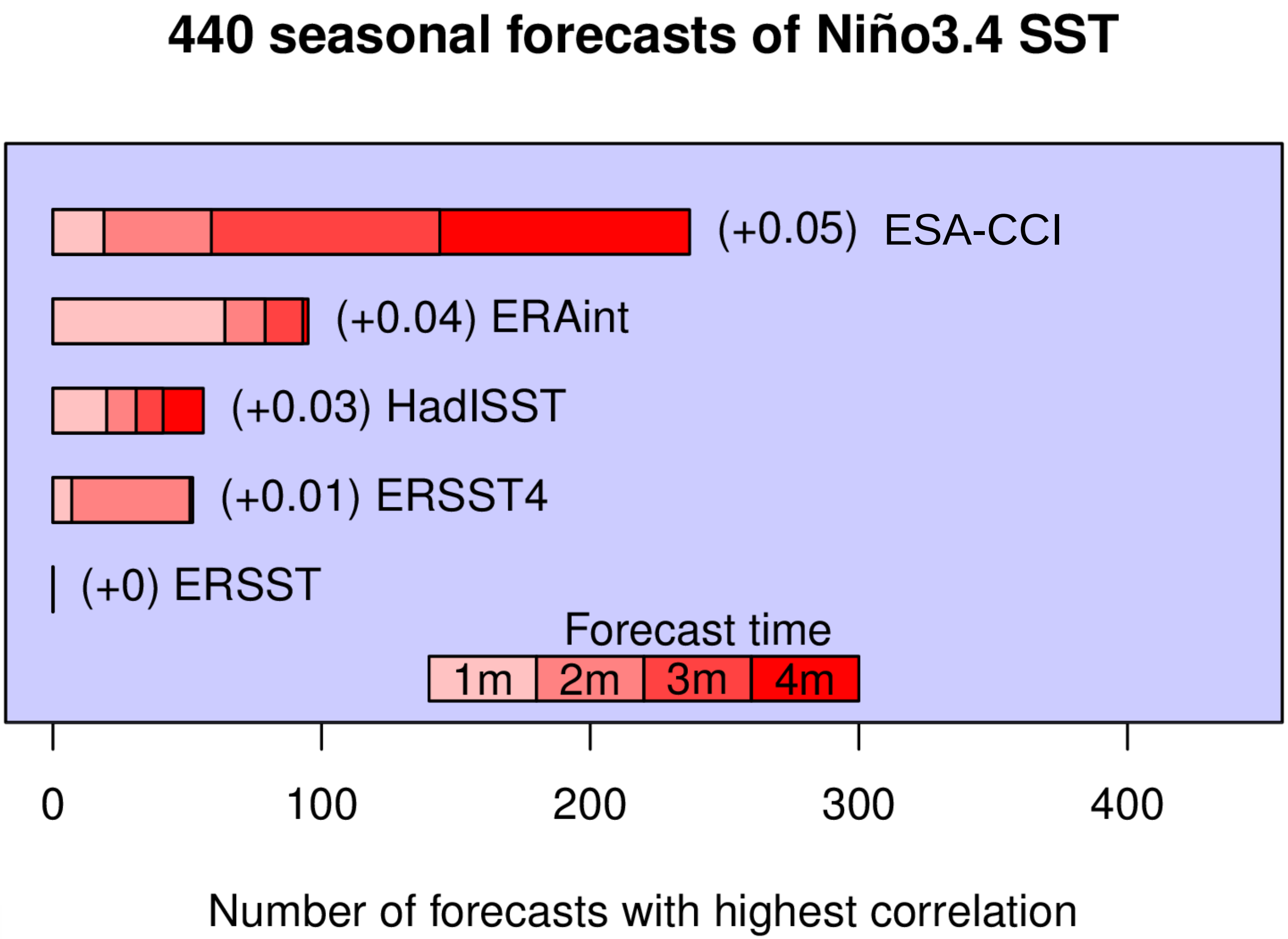


Correlation when using **ERSST** for validation

Correlation when using **ESA** for validation

Differences up to 0.1!

The most advanced product yields on average higher skill to the forecasts



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Such an extreme result is unlikely to have occurred by chance

1. Bootstrapping

- Synthetic data is generated from the known sample covariance matrix of the data that we modify so that, for each forecast, correlations are the same for all observations (our null hypothesis)

- With 10,000 trials, a result as extreme as the one we have happens ~1.2% of the time

Such an extreme result is unlikely to have occurred by chance

1. Bootstrapping

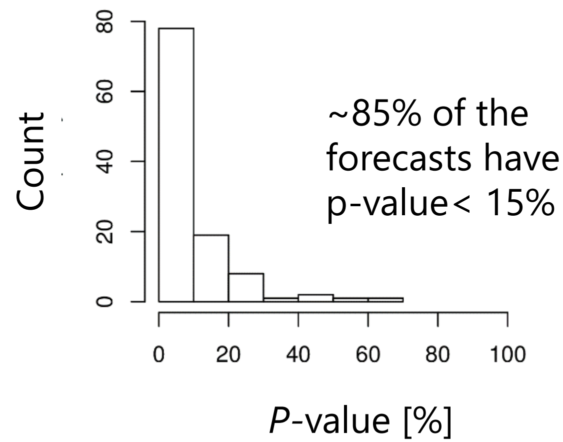
- Synthetic data is generated from the known sample covariance matrix of the data that we modify so that, for each forecast, correlations are the same for all observations (our null hypothesis)

- With 10,000 trials, a result as extreme as the one we have happens ~1.2% of the time

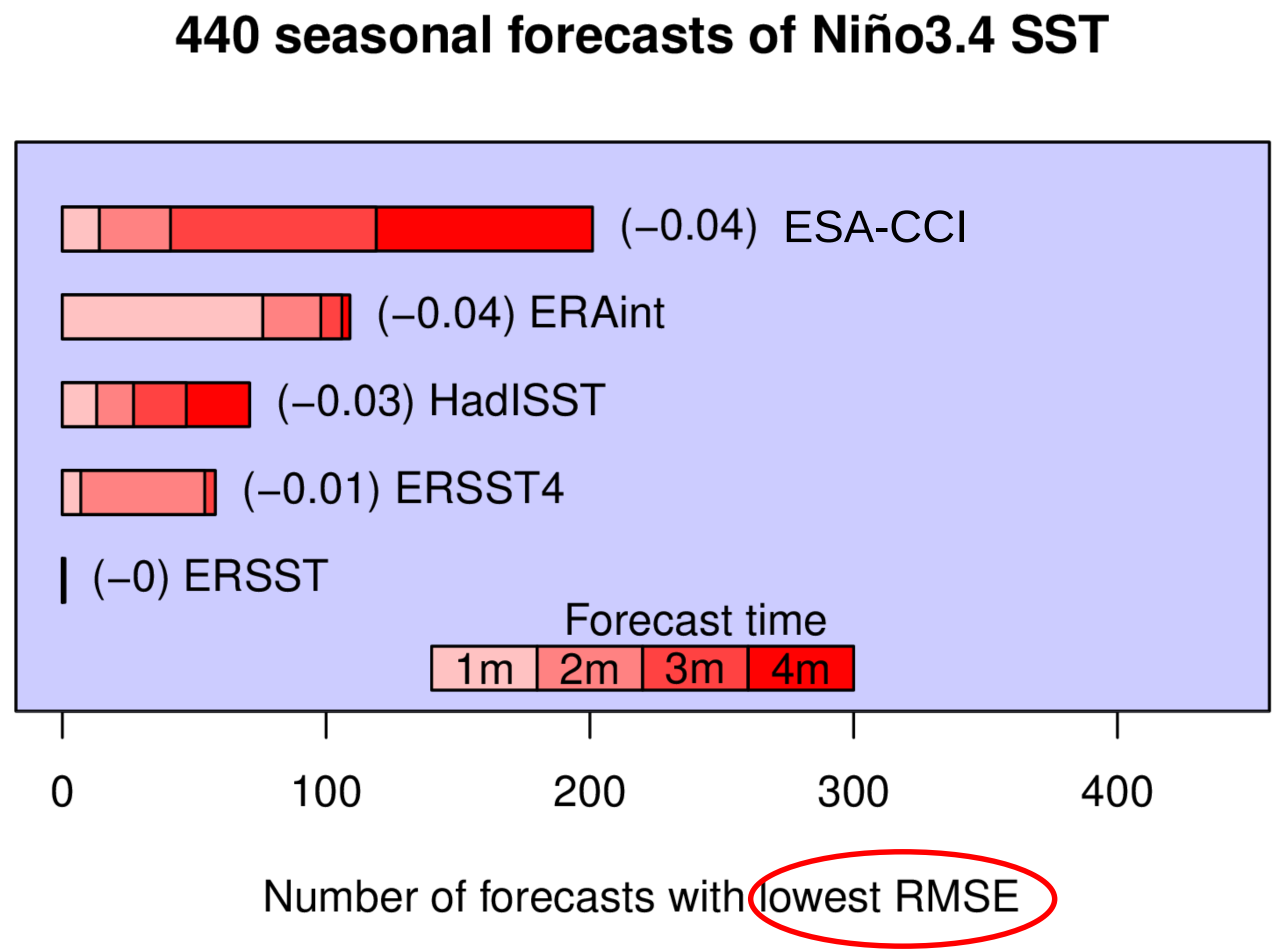
2. Parametric test

[Steiger et al., 1981]
(the test detects changes in correlation in presence of non-independent samples)

p-value of 110 Steiger tests to detect increase of correlation from ERSST (lowest) to ESA (highest)



The result is robust when using an alternative metric of verification: RMSE

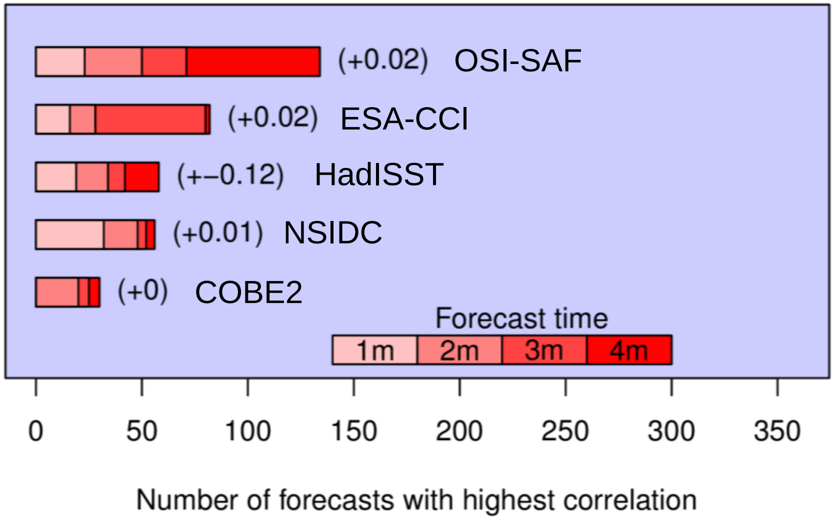


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

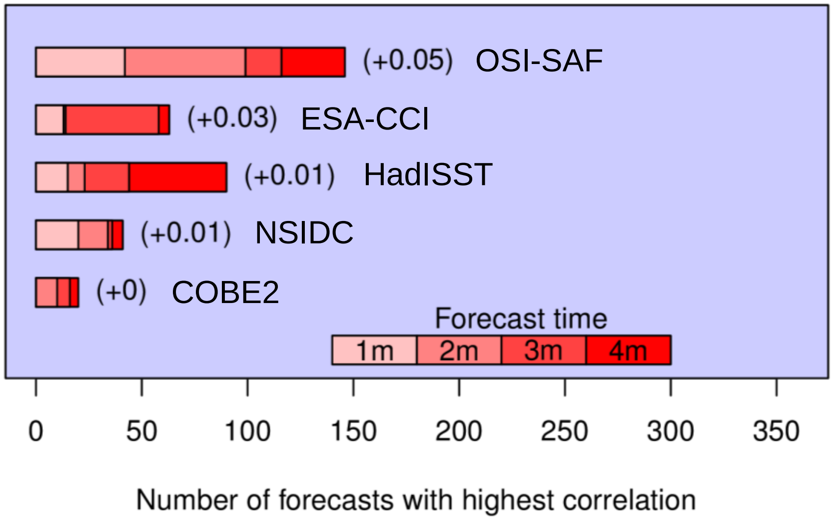
The test was repeated for another test case: sea ice

9 models x 10 members x 4 forecast times = 360 forecasts (ref. 1993-2008)
Initialization month: May

Without detrending



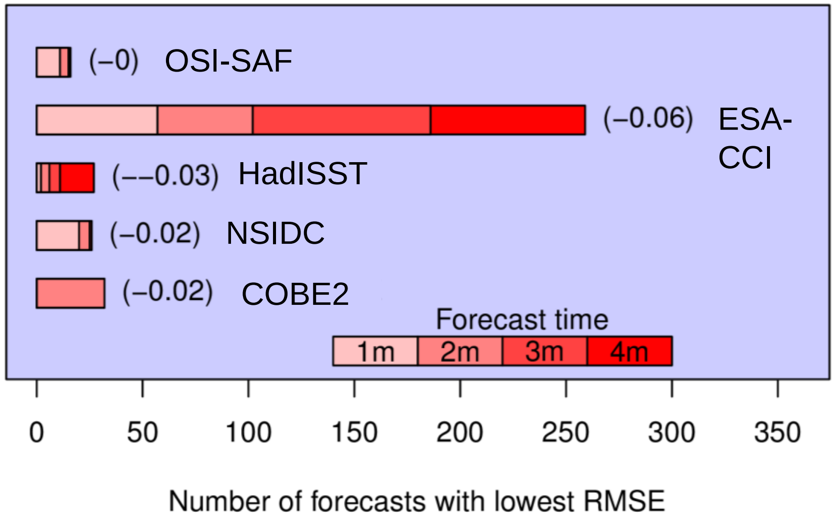
With detrending



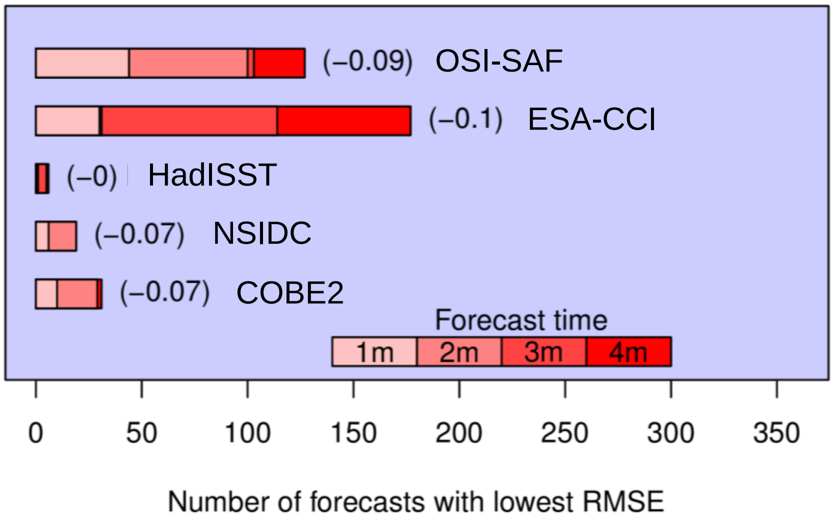
The test was repeated for another test case: sea ice

9 models x 10 members x 4 forecast times = 360 forecasts (ref. 1993-2008)
Initialization month: May

Without detrending



With detrending



Conclusions & Outlooks

Summary | Using three types of arguments (intuitive reasoning, idealized models, real forecasts) we find evidence that better observations increase forecast skill as we estimate it routinely in the community of seasonal forecasting.

Interpretation | These results are best understood if observations and models are considered at the same level (i.e., observations are not superior to models). Observational errors will systematically lower actual forecast skill, in the same way that model errors systematically lower forecast skill.

Recommendations | Models should not always be blamed for low performance. Observations can do this job easily! This overlooked source of error can give differences as large as differences from one model version to another. Modellers should be careful in picking their observational product, or at least use several of them.

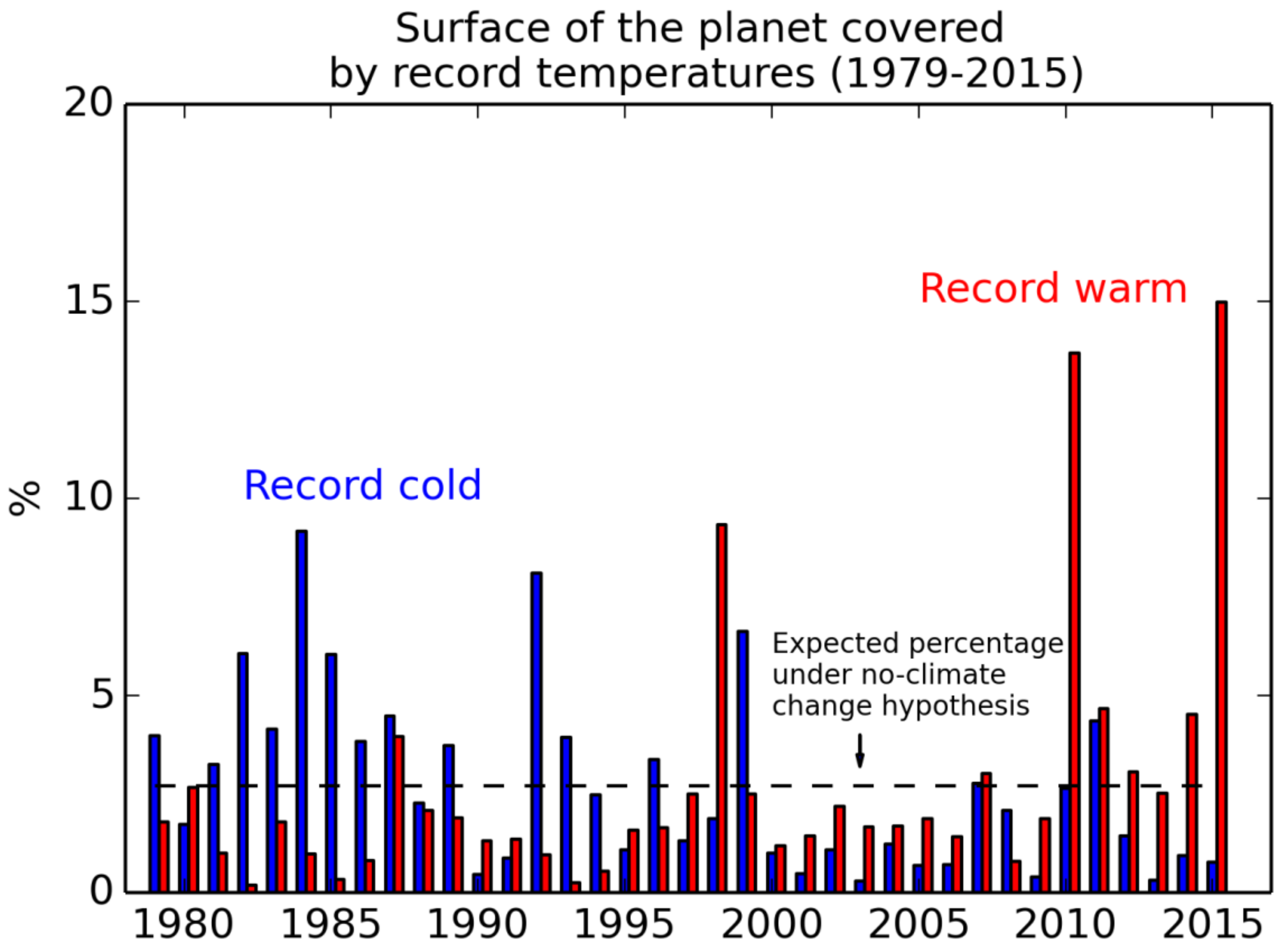
Outlooks | Quantifying observational error propagation over time-averaged periods, space-averaged domains, is key to introduce observational uncertainty in current metrics of performance. Yet these error statistics depend on many unknowns, such as decorrelation time- and space-scales between grid-point, daily error statistics.

Thank you

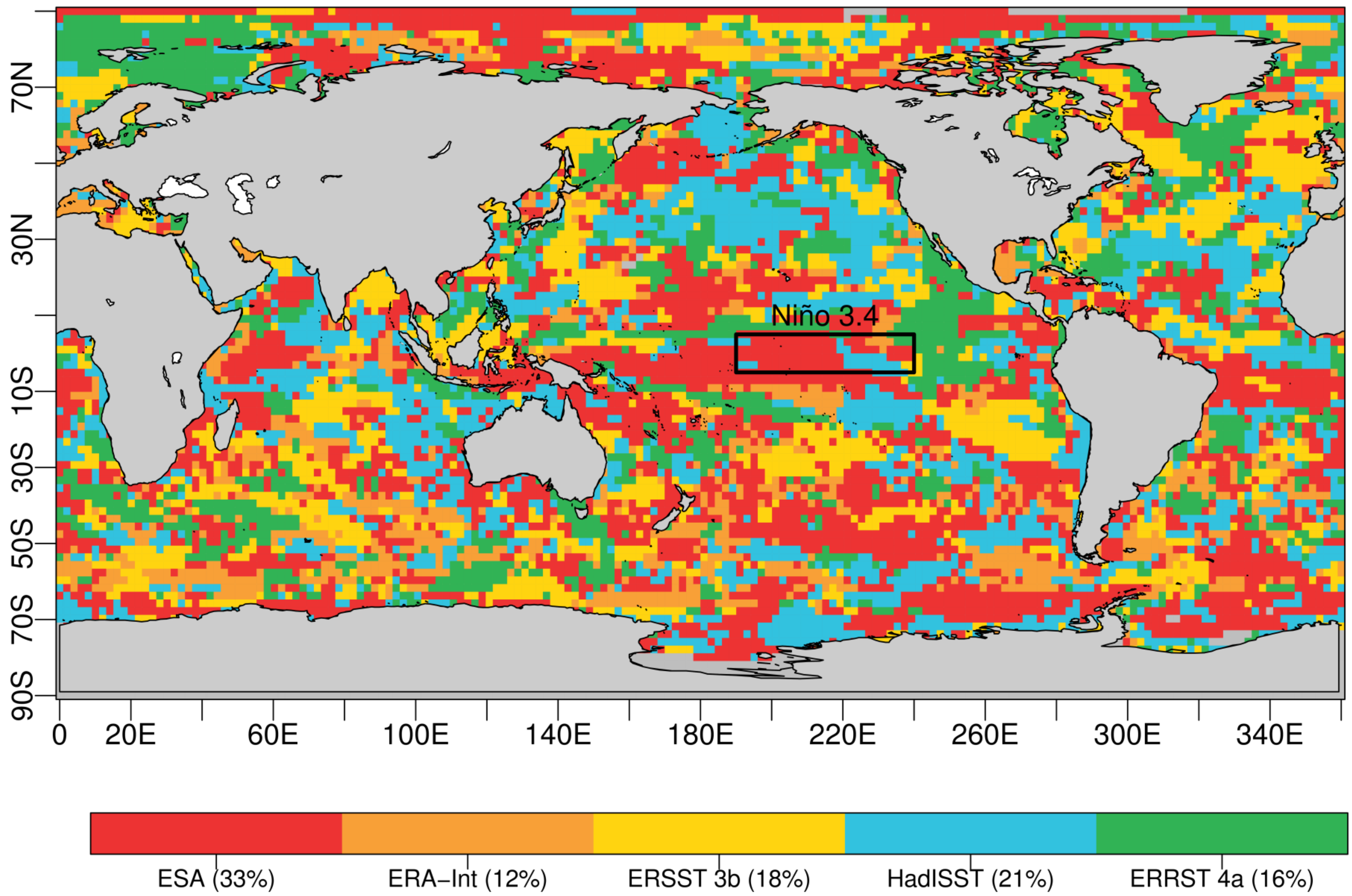
francois.massonnet@bsc.es
@FMassonnet on Twitter
www.climate.be/u/fmasson

http://www.mrwallpaper.com/wallpapers/earth-space-orbit-satellite-1920x1080.jpg

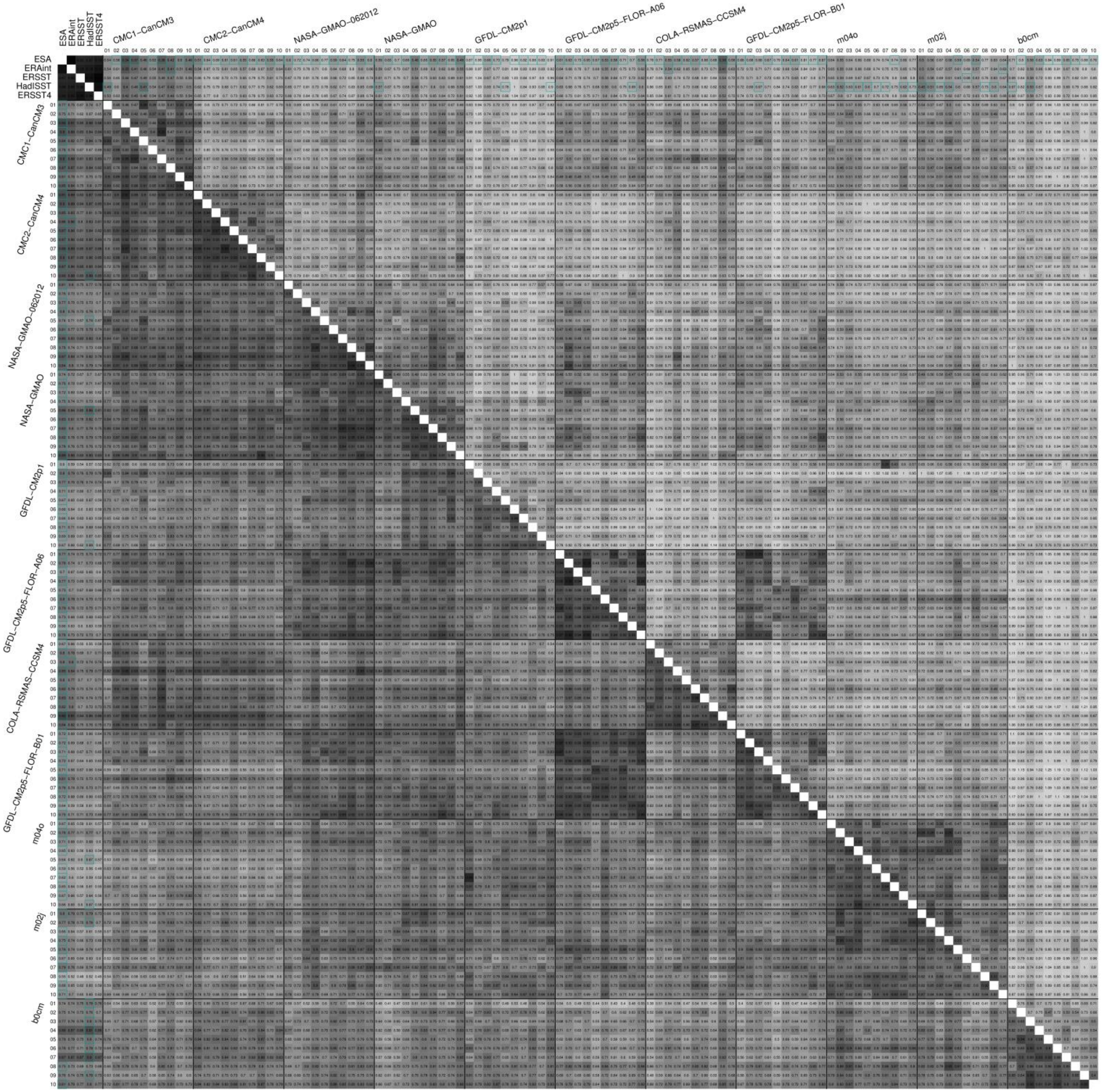
Our climate is changing rapidly

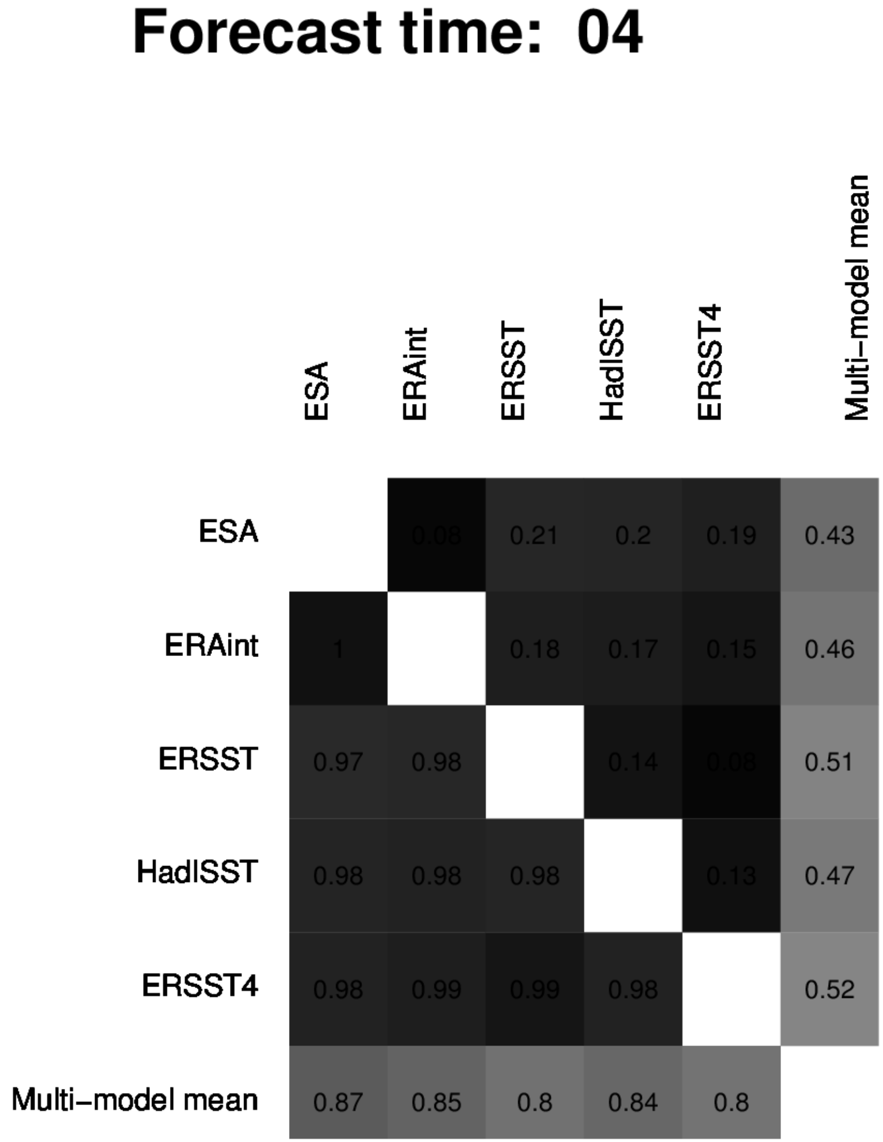


Observation leading to highest correlation in August



Forecast time: 04




Forecast time: 04
ESA
ERAint
ERSST
HadISST
ERSST4
Multi−model mean
0.08
0.21
0.2
0.19
0.43
1
0.18
0.17
0.15
0.46
0.97
0.98
0.14
0.08
0.51
0.98
0.98
0.98
0.13
0.47
0.98
0.99
0.99
0.98
0.52
0.87
0.85
0.8
0.84
0.8

www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Same machine, same climate?

F. Massonnet

M. Asif, M. Acosta, O. Bellprat, E. Exarchou, J. García-Serrano, V. Guemas, M. Ménégoz, O. Mula-Valls, K. Serradell, F. J. Doblas-Reyes, X. Yepes

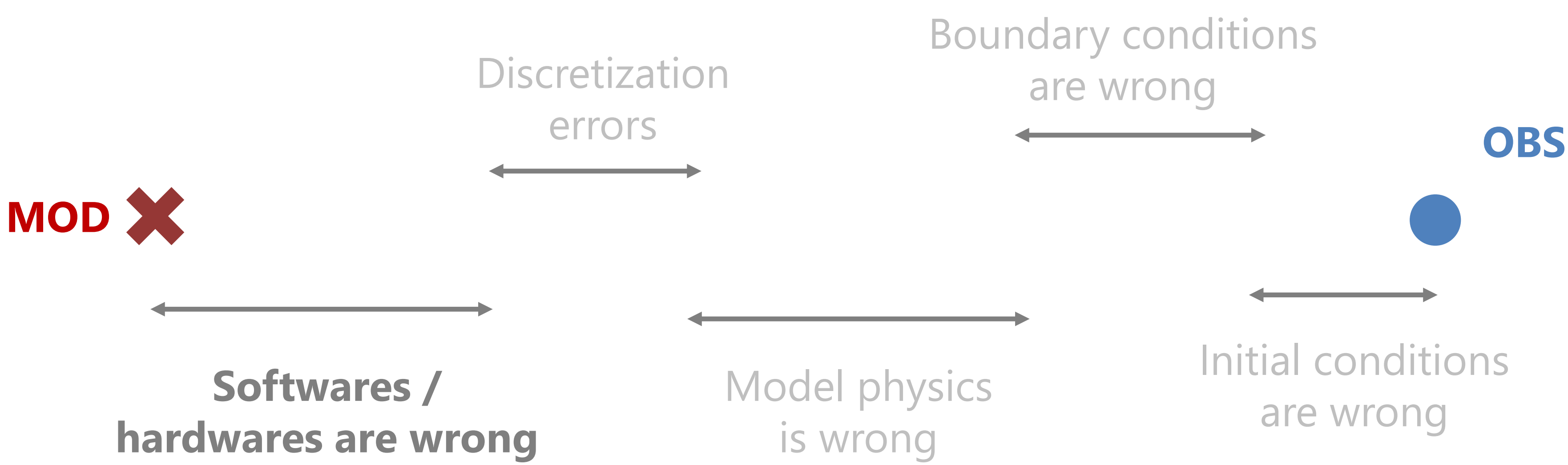
ECMWF
29th February 2016

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

SPECS
Seasonal-to-decadal climate Prediction for the improvement of European Climate Services

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Identifying sources of model error



Software / harwdare is a multiple and underestimated source of model error

Round-off errors and floating-point representation

The order matters: associativity is no longer valid

$$(\sqrt{2}.\pi).e = \sqrt{2}.(\pi.e)$$

12.0771 ≠ 12.0777

Agressive optimization

Can degrade accuracy [Thomas et al. Wea. And Forecast., 2002]

Number of processors and their distribution

Processor topology defines order of operations [Thomas et al., Wea. And Forecast., 2002; Senoner et al., AIAA, 2008]

Compiler version

Different FORTRAN compilers can produce different outcomes [Lawrence et al., EOS, 1999]

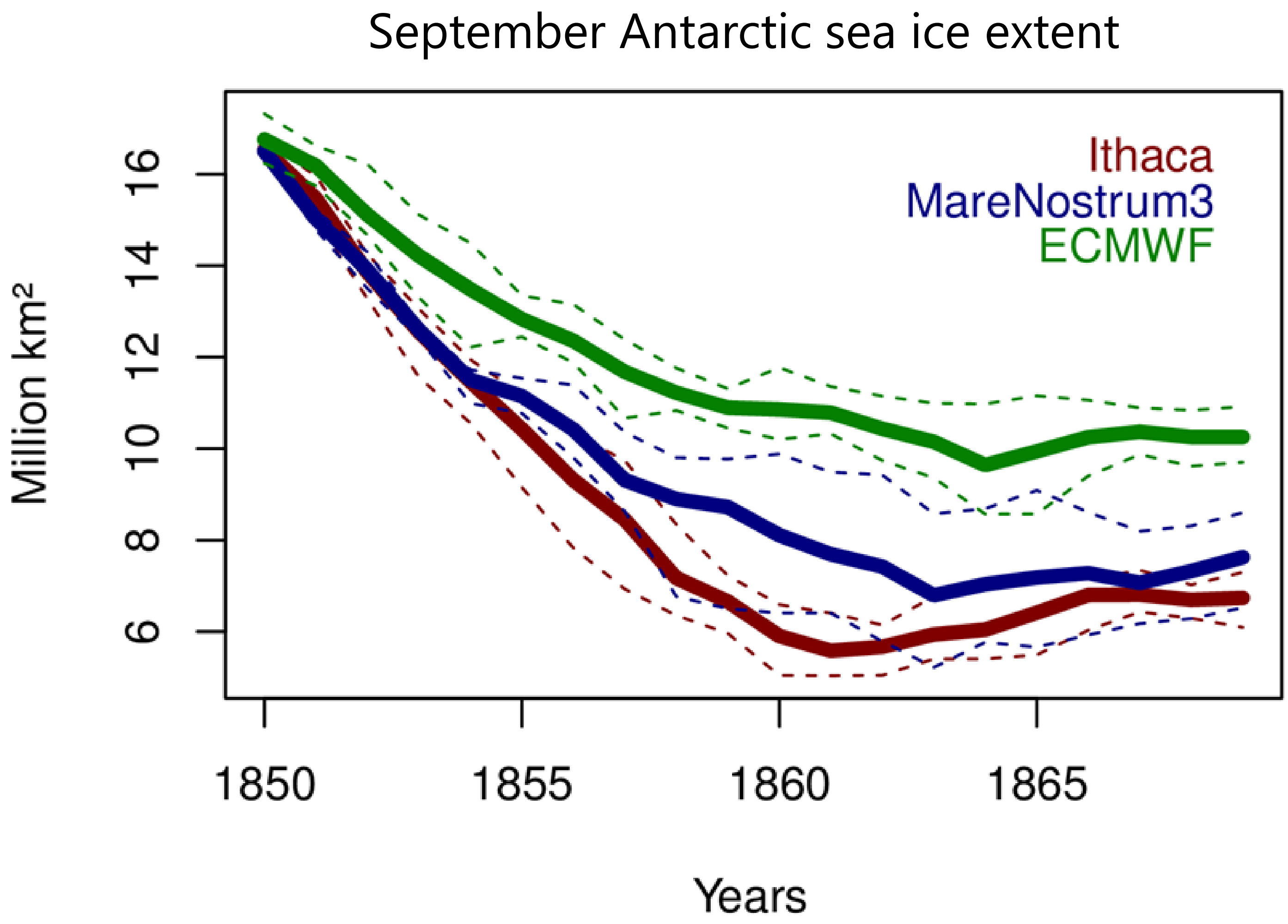
Unpredictable hardware failures

[Düben and Palmer, Mon. Wea. Rev., 2014]

	Machine 1 (Mare Nostrum, BSC)	Machine 2 (ECMWF)	Machine 3 (Ithaca, CFU)
Motherboard	IBM	CRAY	ORACLE
Operating system	SUSE	CRAY LINUX environment	SUSE
Compilation flags	Identical	Identical	Identical
NetCDF, GRIB, HDF5 libraries	Different	Different	Different
# of processors	22+384+96	22+480+96	22+32+16

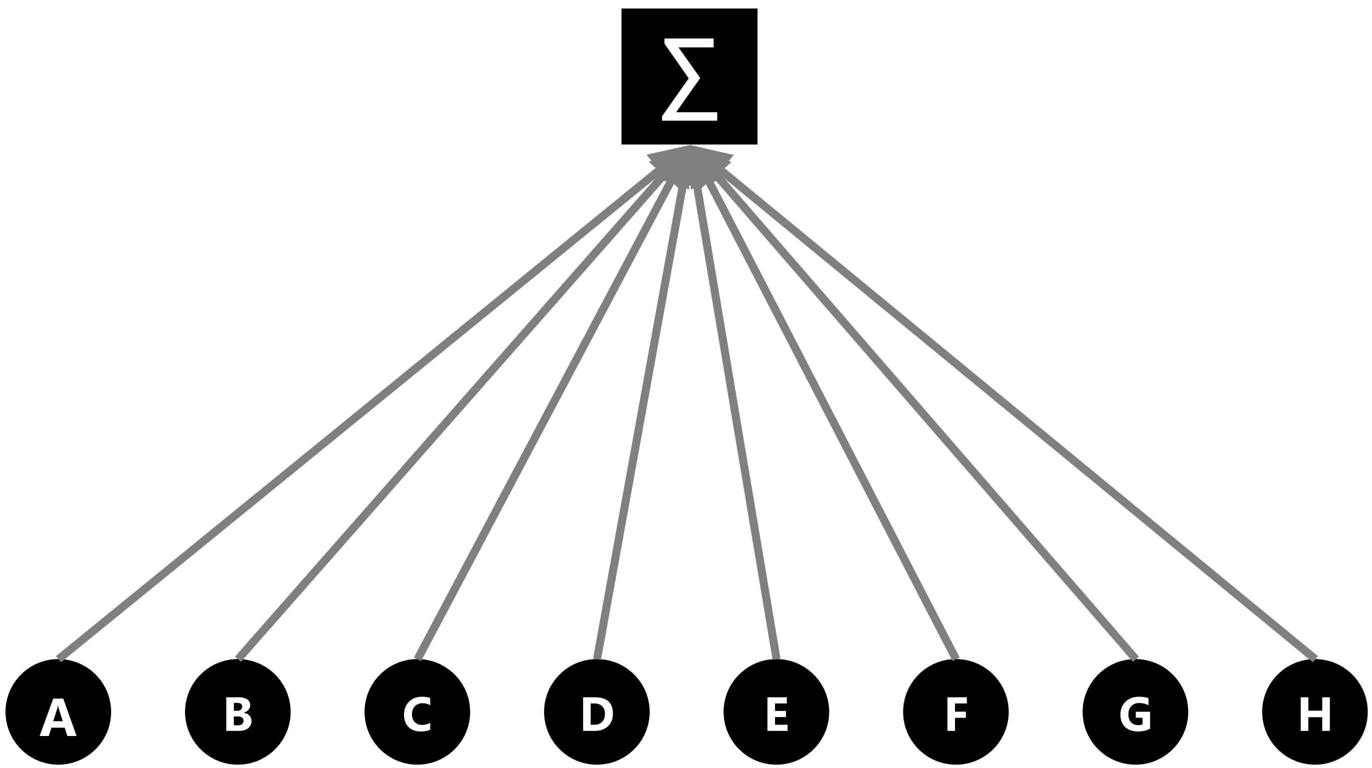
`Autosubmit` ensures identical configurations

Initial results showed that hardware/software *has* an influence on the simulated climate



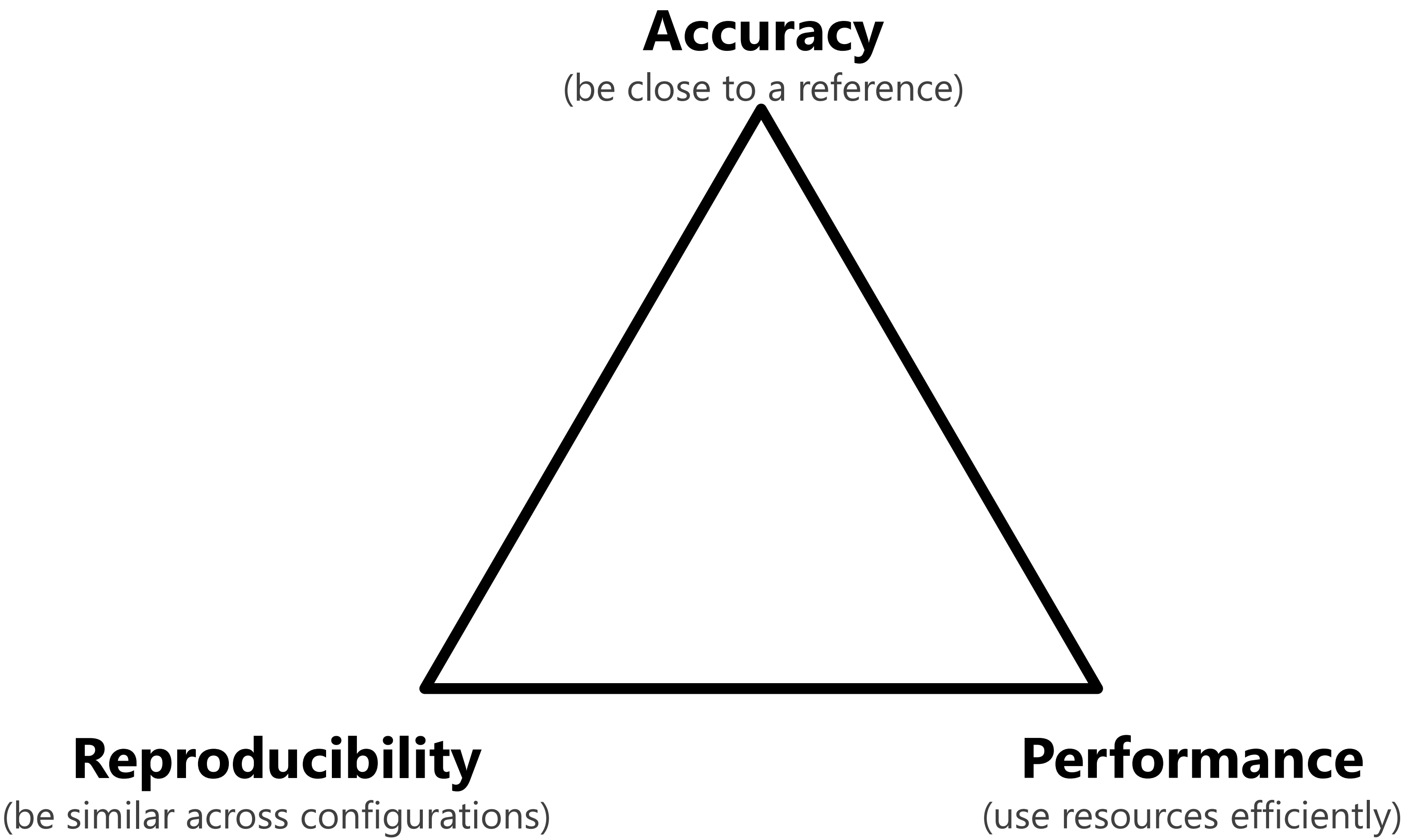
Non-deterministic task in parallel applications

MPI reduction(+): order of summation depends on several external factors



8-cores MPI job

Model development has the following objectives



Compiler options let you control the tradeoffs between accuracy, reproducibility and performance

www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Topics to discuss

F. Massonnet

ECMWF
29th February 2016

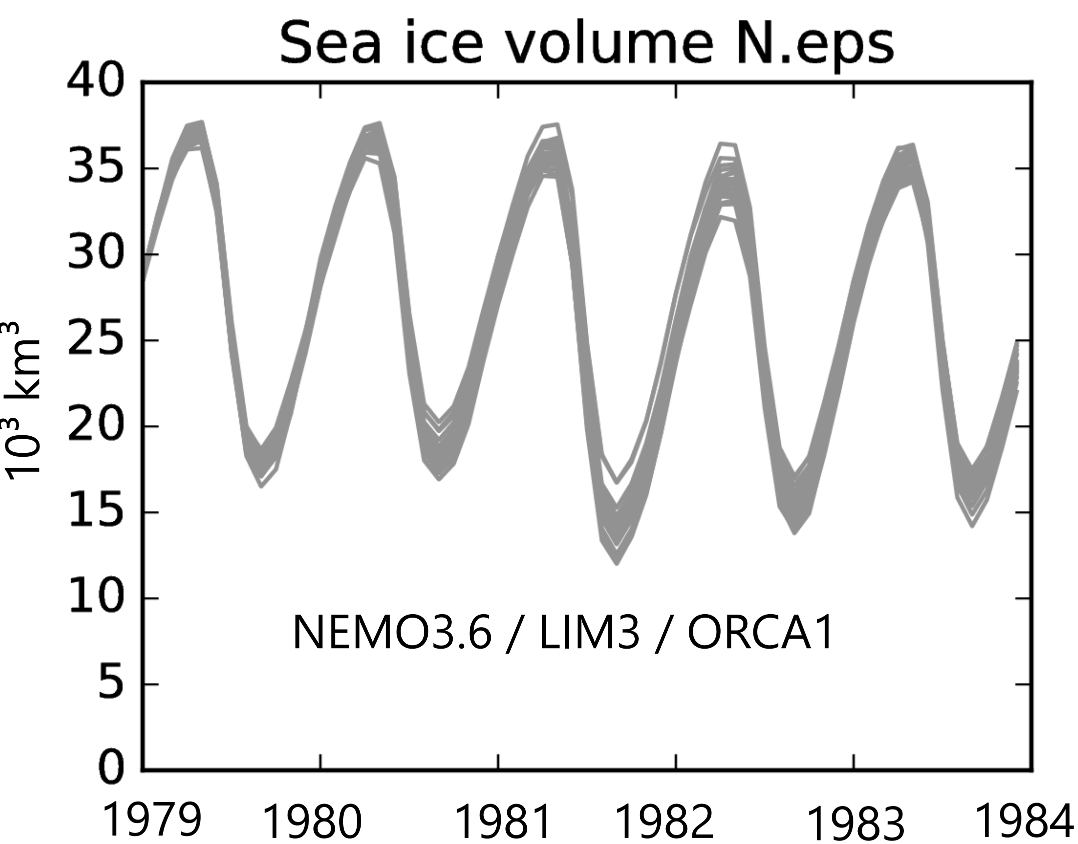
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

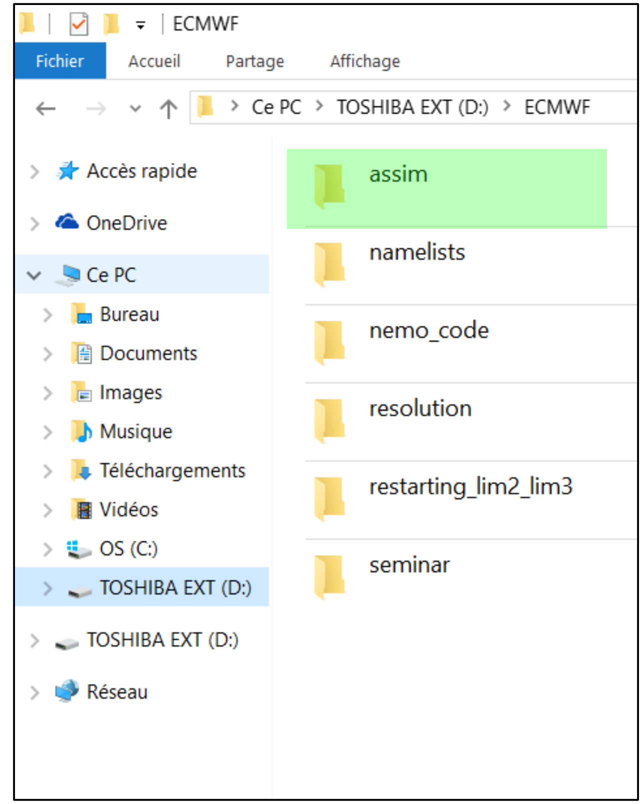
SPECS
Seasonal-to-decadal climate Prediction for the improvement of European Climate Services

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Topics to discuss

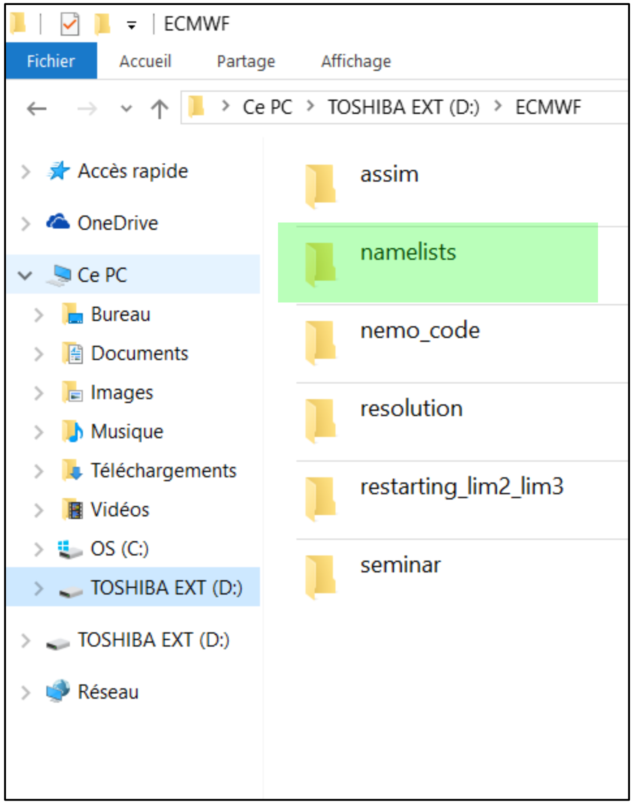
25-member ensemble ocean-sea ice simulation (1979-1982)





- Perturbed atmospheric forcing
- Ready for data assimilation? SMOS?
- Ocean/ice covariances
- …

Topics to discuss



I have our NEMO3.6 namelists for three resolutions (ORCA2, 1, ¼°)

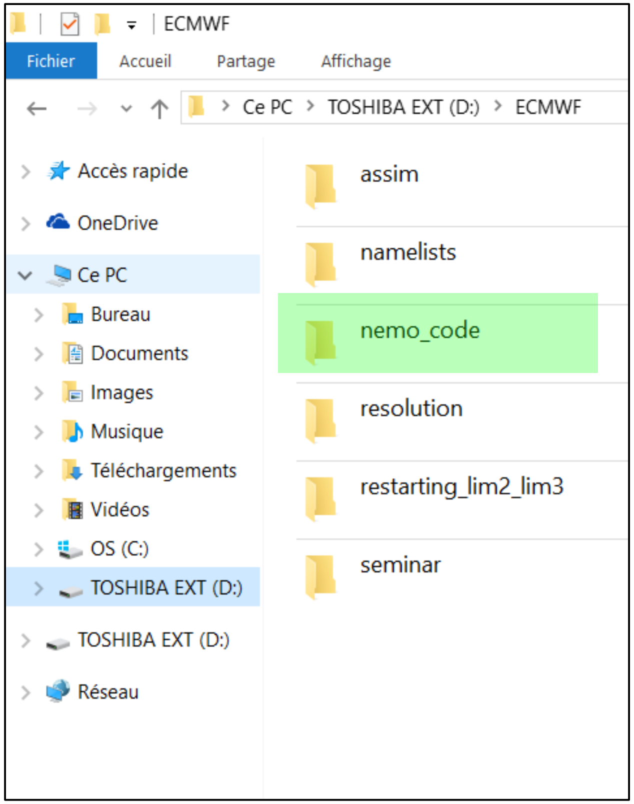
I also can point to relevant configuration files (bathymetry, runoff, initial conditions, etc.) to make sure we will be able to exchange restarts later on.

Topics to discuss

I have the NEMO code that was used to make the latest EC-Earth 3.2 simulationst at BSC.

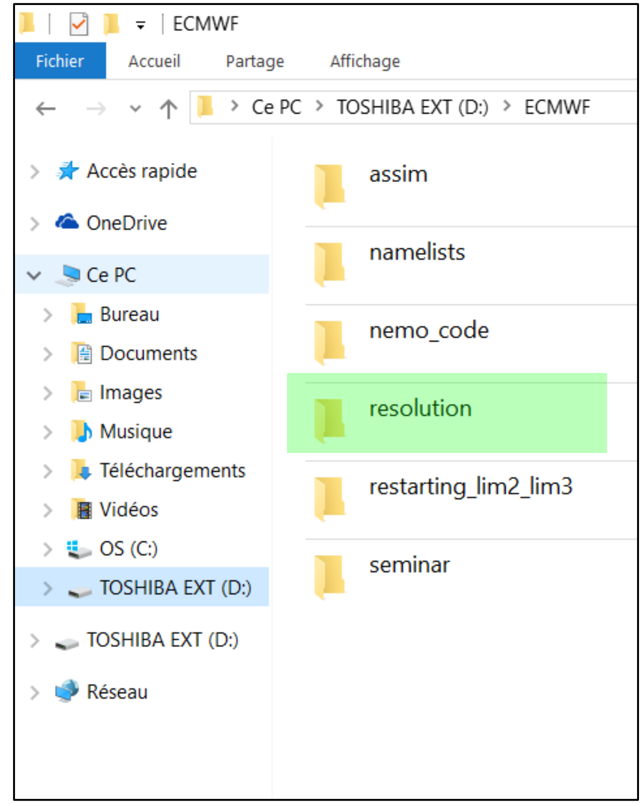
We can discuss coupling strategy, although I'm not an expert in that.

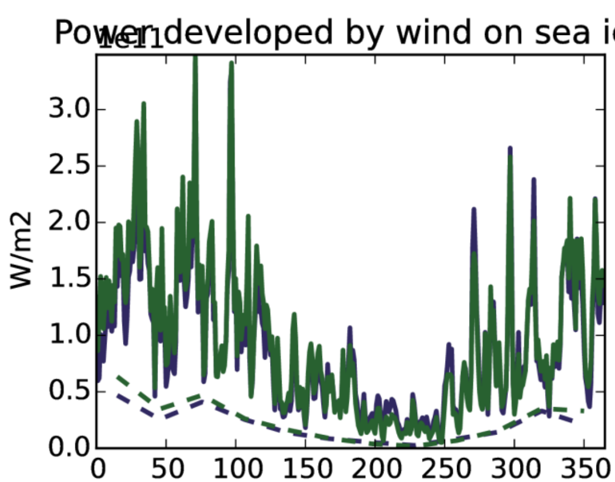
I'm also happy to make tests regarding the importance of salinity in sea ice. Simulations can be launched from here and sensitivity experiments carried on.

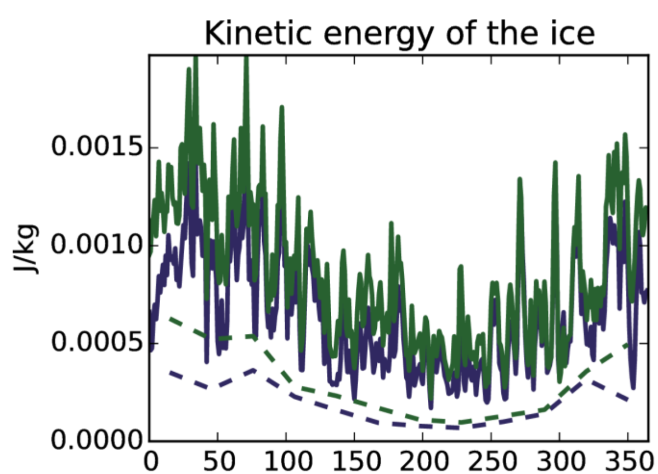


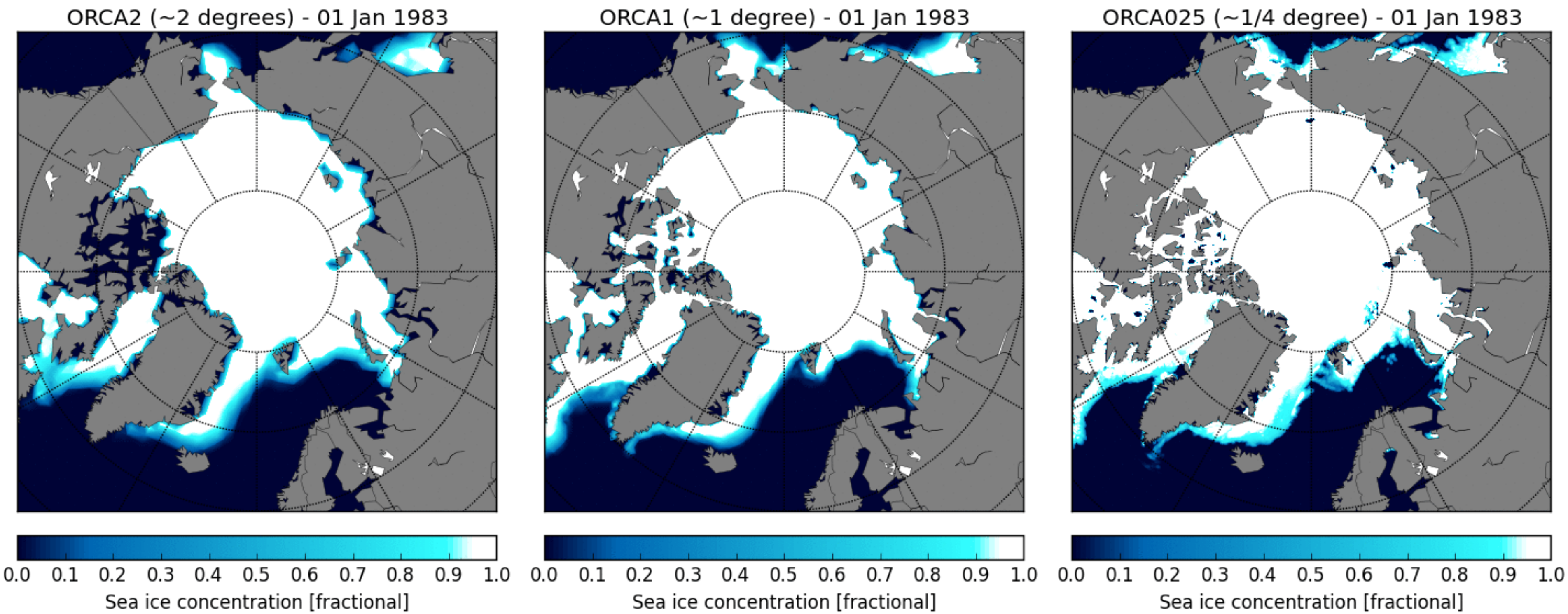
Topics to discuss

I have a set of three simulations at ORCA2, ORCA1, ORCA1/4°, and am happy to discuss how resolution can affect physics.

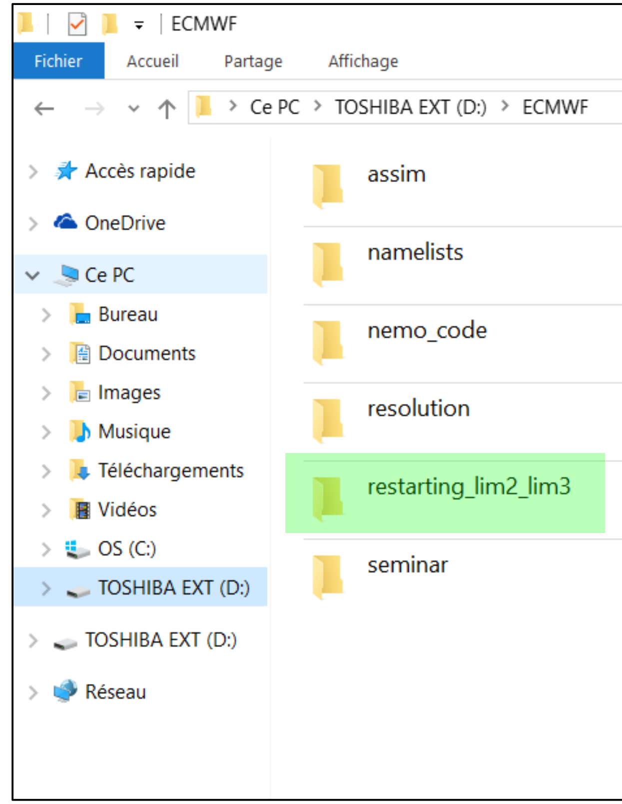







ORCA2 (~2 degrees) - 01 Jan 1983
ORCA1 (~1 degree) - 01 Jan 1983
ORCA025 (~1/4 degree) - 01 Jan 1983
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Sea ice concentration [fractional]
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Sea ice concentration [fractional]
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Sea ice concentration [fractional]

Topics to discuss



1. Let the EnKF analyze *as is* each variable following the determistic framework of Sakov and Oke (2008). In particular, sea ice volume and area in each ice thickness category is updated at this step.
2. Reset prognostic variables to zero if the analyzed total concentration is less than zero.
3. Update the sea ice and snow content in each layer of each category proportionally to the corresponding volume changes in the same layer and category. (The snow and sea ice heat contents are not updated by the EnKF itself).
4. Rebin ice categories: ice and snow areas, volumes and heat contents are transferred to neighbouring categories in case the thickness exceeds the limits.
5. Shrink the ice: if total concentration exceeds 1, spread the excess over each category proportionally to each category's area.
6. Rebin ice categories (same as 4.), as sea ice thickness may have changed in 5.
7. Reset sea surface temperature to freezing point of seawater if sea surface temperature is below the freezing point; the freezing point of seawater in NEMO is a function of sea surface salinity.

From my PhD thesis

`sanity_check.F90`

`sanity_check.py`